The use of proteomics studies in identifying moonlighting proteins

Constance Jeffery [1]*

[1] Department of Biological Sciences, University of Illinois at Chicago, 900 S Ashland Ave, Chicago, IL 60607, USA

* **Corresponding author:** cjeffery@uic.edu, Tel: +1-312-996-3168, Fax: +1-312-413-2691

Running Head: Moonlighting Proteins in Proteomics

i. Abstract

Proteomics studies that characterize hundreds or thousands of proteins in parallel can play an important part in the identification of moonlighting proteins, proteins that perform two or more distinct and physiologically relevant biochemical or biophysical functions. Functional assays, including ligand binding assays, can find a surprising second function for a protein that was previously identified as performing a different function, for example, a DNA binding ability for an enzyme in amino acid metabolism. The results of large scale assays of protein-protein interactions, gene knockouts, or subcellular protein localizations, or bioinformatics analysis of amino acid sequences and three-dimensional structures, can also be used to predict that a protein has additional functions, but in these cases it is important to use biochemical and biophysical methods to confirm the protein can perform each function.

ii. Key Words

moonlighting proteins, multifunctional proteins, protein function prediction, proteomics

1. Introduction

The goal of many proteomics studies is to identify protein functions. Complicating this task is the ability of a single protein to have different functions in different cellular processes, with different ligands or different protein partners, in different cell types, and/or in different subcellular locations. Hundreds of proteins have been identified as these moonlighting proteins, which comprise a subset of multifunctional proteins that perform two or more distinct and physiologically relevant biochemical or biophysical functions that are not due to gene fusions, multiple RNA splice variants, or pleiotropic effects (1). Some of the first moonlighting proteins to be identified were taxon-specific crystallins (2,3), proteins that are found in high concentration in the lens of the eye but function as enzymes in other cell types. For example, zeta-crystallin from the guinea pig lens is identical to the enzyme quinone oxidoreductase (4). Over 300 moonlighting proteins are described in the online MoonProt Database (5). As a group, the known moonlighting proteins perform a large variety of functions and combinations of functions and don't share sequence or structural motifs or other physical characteristics that enable easy identification. Although interpreting the results of proteomics projects might be complicated by the presence of moonlighting proteins, the diversity of moonlighting proteins means that these large-scale projects to characterize proteins, without a prior hypothesis about each protein's function, can be the best way for finding more moonlighting proteins and their multiple functions.

3. Methods

Proteomics experiments can be used to help identify proteins with multiple functions in both direct and indirect ways. Projects based on a functional assay can identify a second function for

a protein that already has a known function. Projects that test other protein characteristics can also be used to suggest that some proteins have a second function, although they might or might not provide information about what the function is. Bioinformatics analyses used alone or in combination with proteomics projects can be used to suggest which other proteins might also have multiple functions.

I. Experimental Methods

The proteomics methods that have been the most useful in finding proteins with multiple functions include those that test binding to a specific molecule, protein-protein interactions, the results of gene knockout experiments, and cellular localization.

Binding studies

Proteomics studies that involve screening hundreds or thousands of proteins to find those that bind to DNA, extracellular matrix, or other macromolecules have identified several dozen proteins that were already known to have a different function. This is not too surprising because dozens of the known moonlighting proteins have at least one function that involves binding to another molecule - as a cell-surface receptor for a soluble ligand or the extracellular matrix, as a secreted ligand binding to a receptor on another cell, or as a DNA or RNA binding protein.

The use of microarrays of proteins, DNA oligonucleotides, or RNA oligonucleotides enables screening of vast numbers of proteins to find those that bind to a chosen macromolecule. Hall and coworkers screened a microarray of yeast proteins for binding to DNA oligonucleotides and identified mitochondrial Arg5,6 (N-Acetylglutamate kinase /N-acetylglutamyl-

phosphate reductase), an enzyme in the arginine biosynthetic pathway, as having DNA binding activity. Complementary chromatin immunoprecipitation experiments and gene deletion experiments confirmed that Arg5,6 is a transcription regulator for several specific nuclear and mitochondrial genes (6).

Assays to identify proteins that bind to a specific protein also identified several dozen yeast and bacterial proteins that perform one function when expressed inside the cell and a second function when displayed on the cell surface. A proteomic study of cell wall proteins from the pathogenic fungus *Candida albicans* was used to identify eight cytosolic proteins (phosphoglycerate mutase, alcohol dehydrogenase, thioredoxin peroxidase, catalase, transcription elongation factor, glyceraldehyde-3-phosphate dehydrogenase, phosphoglycerate kinase and fructose bisphosphate aldolase) as cell surface receptors for host plasminogen (7). A similar study of the intestinal "pro-biotic" bacterium *Bifidobacterium lactis* identified the cytosolic enzymes bile salt hydrolase, glutamine synthetase, and phosphoglycerate mutase and the chaperone DnaK also to bind host plasminogen when expressed on the cell surface (8).

Protein-protein interactions

Proteomics-scale studies of protein-protein interactions, such as yeast 2-hybrid assays, often yield results that are more complex than expected, with a single protein being found to interact with proteins acting in multiple biochemical pathways, molecular machines or multiprotein complexes. These results are sometimes interpreted as being due to false positives, but interacting with multiple groups of proteins from different cellular processes is a common characteristic of moonlighting proteins, and these protein-protein interaction results could be due

to physiologically relevant interactions (9). In a study of human proteins, Chapple and coworkers combined protein-protein interaction information with analysis of the protein functional annotations to identify 430 proteins that they described as being extreme multifunctional proteins (10). Further analysis of protein-protein interaction networks from humans and other species could lead to predictions of additional proteins that might interact with different groups of protein partners to perform different functions, along with suggestions of the types of additional functions based on the identities of the interacting proteins. Follow-up testing through biochemical or biophysical assays would be needed to confirm the observed interactions are due to a second function and not due to false positives, or to proteins that interact with multiple proteins as part of a single function, such as in a signaling pathway, or proteins that perform the same function in different cellular locations.

Gene knockouts

When a single protein participates in several cellular processes, deletion of the gene encoding the protein can result in phenotypes that are more complex than can be explained by the loss of a single function. Several labs made use of yeast genetics to look for enzymes for which replacement of the wild type enzyme with a catalytic deficient mutant does not recapitulate the results of the complete gene knockout. Because the mutant protein exhibits only part of the deletion phenotype, the wild type protein must have a second function. The *S. cerevisiae* Bat2 transaminase in sugar and amino acid metabolism and the isoleucine/valine biosynthetic enzymes Ilv1 and Ilv2 were found to have second functions in addition to their respective catalytic functions (11). The Alt1 alanine transaminases from two other yeast species, *Lacchancea kluyveri* (LkAlt1) and *Kluyveromyces lactis* (KlAlt1), were also found to be moonlighting proteins (12).

Expression patterns/Cellular Localization

Proteomics projects that determine a protein's cellular location(s) can be used to suggest that it might have multiple functions. Many of the known moonlighting proteins perform their different functions in different subcellular locations or cell types. For example, several dozen cytosolic proteins, like the plasminogen binding proteins mentioned above, have a second function as a receptor or adhesin on the cell surface in bacteria, humans, and many other species. Glyceraldehyde 3-phosphate dehydrogenase (GAPDH) was the first cytoplasmic protein found to be attached to the surface of pathogenic streptococci (13), and several dozen cytosolic enzymes in multiple species have also been found to be displayed on the cell surface, where they play roles in signaling, adhesion, or acquiring nutrients. Studies to identify all the proteins on the cell surface of dozens of bacterial species, through cell fractionation and isolation of proteins followed by identification through mass spectrometry, found that many other cytoplasmic proteins are also attached to the cell surface, and some may be additional moonlighting proteins (14,15). Other methods of studying protein localization can also be scaled up and used in proteomics studies. In a recent study using antibody-based immunofluorescence of 12,003 human proteins in 30 subcellular structures and 13 organelles, about half of the proteins were found in more than one compartment and might also include candidates for moonlighting proteins (16).

The taxon-specific crystallins mentioned above clearly have two functions because these enzymes are found in a high concentration in the lens of the eye where their catalytic substrates are not found, and the known intracellular/cell surface moonlighting proteins have been tested

through binding studies to confirm the presence of a second function. In other cases, finding a protein in a subcellular location where it is not expected to perform its known function can suggest that the protein has a second function, but experimental evidence that the protein is performing a different function in each location is needed to confirm that the protein is multifunctional. For example, proteins that move between cellular compartments as part of a signaling pathway would not be considered moonlighting proteins.

II. Bioinformatics Analysis

The lack of common sequence or structural characteristics among the moonlighting proteins has made it difficult to develop a universal computational method to predict that a protein has more than one function, but several labs are developing bioinformatics methods by using collections of known moonlighting proteins, such as the MoonProt Database, as a positive control set. Unfortunately, a true negative control set of proteins that don't have multiple functions is not available because it's currently not possible to know if a protein has only one function or if it has additional functions that have not yet been identified.

Large-scale searches of the literature and database annotation, including searches for proteins with diverse GO terms in UniProt (17,18) have shown some success in identifying proteins with multiple functions in diverse processes (19, 20). Searches for amino acid sequence or structural motifs known to be associated with specific protein functions can help identify proteins that have motifs corresponding to multiple functions. For example, the use of an X-ray crystal structure revealed the *Streptomyces coelicolor* albaflavenone synthase also has a terpene synthase active site (21). However, the use of motifs as a tool for prediction of function is limited by the lack of known motifs for many classes of functions, for example, protein-protein interactions. As more

information becomes available about sequence motifs, protein-protein interaction surfaces, the constellations of amino acids that make up catalytic sites, and the three-dimensional structures of moonlighting proteins, the ability of these methods to find the multiple functions of moonlighting proteins might be improved (22, 23).

Some recent methods combine analysis of protein sequences and structures with information about the results from proteomics projects (protein-protein interactions, cellular locations, etc.) (24, 25). Because of the challenges in identifying moonlighting proteins, these combined methods might have the most success in the future.

Conclusions

The presence of moonlighting proteins adds to the complexity of the results of proteomics studies, but these large-scale methods are valuable for identifying more examples of these proteins, which are otherwise usually found through serendipity. One caution, however, is that only the proteomics methods that directly test for a specific function, such as the plasminogen binding assays mentioned above, provide evidence that the protein performs that function. The results of the other proteomics and bioinformatics methods described herein can usually only be used to predict that a protein has a second function. It is necessary to use biochemical and biophysical methods to confirm that a protein performs both functions.

References

1. Jeffery CJ (1999) Moonlighting proteins. Trends Biochem Sci. 24(1):8-11. PMID: 10087914

2. Piatigorsky J, Wistow GJ (1989) Enzyme/crystallins: gene sharing as an evolutionary strategy. Cell 57:197–199

3. Wistow GJ, Kim H (1991) Lens protein expression in mammals: taxon specificity and the recruitment of crystallins. J Mol Evol 32:262–269.

4. Huang QL, Russell P, Stone SH, Zigler JS Jr. (1987) Zeta-crystallin, a novel lens protein from the guinea pig. Curr Eye Res 6:725-32. PMID: 3595182

5. Mani M, Chen C, Amblee V, Liu H, Mathur T, Zwicke G, Zabad S, Patel B, Thakkar J, Jeffery CJ (2015) MoonProt: a database for proteins that are known to moonlight. Nucleic Acids Res 43:D277–D282

6. Hall DA, Zhu H, Zhu X, Royce T, Gerstein M, Snyder M (2004) Regulation of gene expression by a metabolic enzyme. Science 306:482-4

7. Crowe JD, Sievwright IK, Auld GC, Moore NR, Gow NA, Booth NA (2003) *Candida albicans* binds human plasminogen: identification of eight plasminogen-binding proteins. Mol Microbiol 47:1637-51. PMID: 12622818

8. Candela M, Bergmann S, Vici M, Vitali B, Turroni S, Eikmanns BJ, Hammerschmidt S, Brigidi P (2007) Binding of human plasminogen to Bifidobacterium. J Bacteriol 189:5929-36 doi: 10.1128/JB.00159-07

9. Gómez A, Hernández S, Amela I, Piñol J, Cedano J, Querol E (2011) Do protein-protein interaction databases identify moonlighting proteins? Mol Biosyst 7:2379-82. doi: 10.1039/c1mb05180f

10. Chapple CE, Robisson B, Spinelli L, Guien C, Becker E, Brun C (2015) Extreme multifunctional proteins identified from a human protein interaction network. Nat Commun 6:7412. doi:10.1038/ncomms8412

11. Espinosa-Cantú A, Ascencio D, Herrera-Basurto S, Xu J, Roguev A, Krogan NJ, DeLuna A (2018) Protein Moonlighting Revealed by Noncatalytic Phenotypes of Yeast Enzymes. Genetics 208:419-431. doi: 10.1534/genetics.117.300377

12. Escalera-Fanjul X, Campero-Basaldua C, Colón M1, González J, Márquez D1, González A. (2017) Evolutionary Diversification of Alanine Transaminases in Yeast: Catabolic Specialization and Biosynthetic Redundancy. Front Microbiol 8:1150. doi: 10.3389/fmicb.2017.01150.

13. Pancholi V, Fischetti VA (1992) A Major Surface Protein on Group A Streptococci is a Glyceraldehyde-3-Phosphate-Dehydrogenase with Multiple Binding Activity. J Exp Med 176:415–426.

14. Olaya-Abril A, Jiménez-Munguía I, Gómez-Gascón L, Rodríguez-Ortega MJ (2014) Surfomics: shaving live organisms for a fast proteomic identification of surface proteins. J Proteomics 97:164-76. doi: 10.1016/j.jprot.2013.03.035

15. Wang W, Jeffery CJ (2016) An analysis of surface proteomics results reveals novel candidates for intracellular/surface moonlighting proteins in bacteria. Mol Biosyst 12:1420-1431.

16. Thul PJ, Åkesson L, Wiking M, Mahdessian D, Geladaki A, Ait Blal H, Alm T, Asplund A, Björk L, Breckels LM, Bäckström A, Danielsson F, Fagerberg L, Fall J, Gatto L, Gnann C, Hober S, Hjelmare M, Johansson F, Lee S, Lindskog C, Mulder J, Mulvey CM, Nilsson P, Oksvold P, Rockberg J, Schutten R, Schwenk JM, Sivertsson Å, Sjöstedt E, Skogs M, Stadler C, Sullivan DP, Tegel H, Winsnes C, Zhang C, Zwahlen M, Mardinoglu A, Pontén F, von Feilitzen K, Lilley KS, Uhlén M, Lundberg E (2017) A subcellular map of the human proteome. Science 356: XXX-XXXdoi: 10.1126/science.aal3321.

17. Consortium GO (2015) Gene Ontology Consortium: going forward. Nucleic Acids Res 43:D1049–D1056

18. UniProt Consortium (2015) UniProt: a hub for protein information. Nucleic Acids Res 43:D204–D212

19. Khan IK, Bhuiyan M, Kihara D (2017) DextMP: deep dive into text for predicting moonlighting proteins. Bioinformatics 33:i83-i91. doi: 10.1093/bioinformatics/btx231

20. Pritykin Y, Ghersi D, Singh M (2015) Genome -Wide Detection and Analysis of Multifunctional Genes. PLoS Comput Biol 11:e1004467. doi: 10.1371/journal.pcbi.1004467.

21. Zhao B, Lei L, Vassylyev DG, Lin X, Cane DE, Kelly SL, Yuan H, Lamb DC, Waterman MR (2009) Crystal structure of albaflavenone monooxygenase containing a moonlighting terpene synthase active site. J Biol Chem 284:36711-9. doi: 10.1074/jbc.M109.064683.

22. Khan I, Chitale M, Rayon C, Kihara D (2012) Evaluation of function predictions by PFP, ESG, and PSI-BLAST for moonlighting proteins. BMC Proc 6 Suppl 7:S5. doi: 10.1186/1753-6561-6-S7-S5

23. Hernández S, Franco L, Calvo A, Ferragut G, Hermoso A, Amela I, Gómez A, Querol E, Cedano J (2015) Bioinformatics and Moonlighting Proteins. Front Bioeng Biotechnol 3:90 doi: 10.3389/fbioe.2015.00090.

24. Khan IK, Kihara D (2016) Genome-scale prediction of moonlighting proteins using diverse protein association information. Bioinformatics 32:2281–2288 doi: 10.1093/bioinformatics/btw166

25. Khan I, McGraw J, Kihara D (2017) MPFit: Computational Tool for Predicting Moonlighting Proteins. Methods Mol Biol 1611:45-57. doi: 10.1007/978-1-4939-7015-5_5